SPECIFIC PHONEME INTELLIGIBILITY OF WORDS SUBJECTED TO TIME-SMEARING DISTORTION

by

Russel L. Sergeant and Joseph S. Russotti
Naval Submarine Medical Center
and
Harry S. Cooker
University of Connecticut

SUBMARINE MEDICAL RESEARCH LABORATORY
U.S. NAVAL SUBMARINE MEDICAL CENTER REPORT NO. 649

Bureau of Medicine and Surgery, Navy Department
Research Work Unit M4306.03-2020D.03

Submitted by:

J. Donald Harris

J. Donald Harris, Ph.D.
Head, Auditory Research Branch
Submarine Medical Research Laboratory

Reviewed and Approved by:

Charles F. Gell

Charles F. Gell, M.D., D.Sc. (Med).
Scientific Director
SubMedRschLab

Reviewed and Approved by:

Joseph D. Bloom

J. D. Bloom, CDR, MC, USN
Officer-in-Charge
SubMedRschLab

Approved and Released by:

J. E. Stark

J. E. STARK, CAPT, MC, USN
COMMANDING OFFICER
Naval Submarine Medical Center

Approved for public release; distribution unlimited.

SUMMARY PAGE

THE PROBLEM

One distortion of speech transmitted through the ocean is called "time-smearing". The problem was to evaluate phoneme intelligibility of speech subjected to time-smearing.

FINDINGS

Speech was degraded by computer to simulate various amounts of time-smearing, and further degraded by 4000 and 1200 Hertz low-pass filtering. Results show that the reduction in intelligibility produced by changing the filtering from 4000 to 1200 Hertz was much greater than the reductions caused by altering smear duration. However, as duration increased, intelligibility dropped significantly. The voiceless stop-plosives [p, t] and [k] were the phonemes most severely affected by the time-smearing; the semivowels [r, l] and [w] were least affected.

APPLICATION

Results shown in this report should be considered when constructing special vocabularies for use in communicating by voice from one submarine to another through oceanic conditions where time-smearing effects are present.

ADMINISTRATIVE INFORMATION

This investigation was conducted as part of Bureau of Medicine and Surgery Work Unit M4306.03-2020D, Sensory Aids for Communications Under Water by Divers and Swimmers, and partially funded by NAVSHIPS Systems Command Task No. 5-23-18-08622 (USL 9-A-611-03-00). It was approved for publication on 25 January 1971, and has been designated as SubMedResLab Report No. 649.

The assistance of Dr. H. Cooker was furnished under ONR Contract with the University of Connecticut (N00014-67-A-0197-0001).

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL CENTER

ABSTRACT

This report concerns a detailed analysis of the intelligibility of spoken consonants in time-smeared speech. "Time-smearing" is the name given to one of the distortions of speech that takes place when it is transmitted through the ocean. Recorded tests of speech materials were treated by computer to simulate varied degrees of time-smearing. In addition, the speech material was further degraded by either a 4000 or 1200 Hertz low-pass filtering. The responses of 320 Navy enlisted men who heard this material were analyzed for phoneme intelligibility. Results show the reduction in intelligibility produced by changing the filter condition was much greater than reductions caused by changing smear duration. As the duration of the smearing increased, intelligibility dropped significantly. The voiceless stop-plosives [p, t] and [k] were the phonemes most severely affected by the time-smear distortion while the semi-vowels [r, l] and [w] were least disturbed. Conclusion: the intelligibility of specific speech sounds should be considered when constructing special vocabularies for transmitting speech signals through the ocean, if time-smearing effects are present.

SPECIFIC PHONEME INTELLIGIBILITY OF WORDS SUBJECTED TO TIME-SMEARING DISTORTION

INTRODUCTION

The efficiency of inter-submarine voice communication is affected by certain undesirable changes that occur in the speech signal during its transmission through the ocean. Speech that has been passed through an acoustic channel such as the ocean is distorted by the phenomenon known as "time-smearing" (See Sachs, et al, 1969)[3]. If an underwater source and a somewhat distant receiver are both situated relatively close to the surface of the ocean, the sound radiated from the source may travel toward the receiver over several different pathways. The radiated acoustic energy may be reflected from the ocean floor and the ocean's surface one or more times before it finally reaches the receiver. Since these pathways are of different length, a single burst of acoustic energy from the source arrives at the receiver as a sequence of two or more distinct bursts that are separated in time.

For a particular arrival over any one of the pathways, the signal is still further distorted. Since the ocean's surface may be covered with waves and the ocean floor may be rough, sound will be reflected at the many different facets within the area ensonified by the source. Therefore, after traversing any given pathway, a single burst of acoustic energy will arrive at the receiver as many overlapping bursts. Consequently, a short pulse may be "smeared out" in time. Thus, even in the absence of gross echoes, speech transmitted over such a system may be unintelligible if there is an appreciable amount of "smearing". Furthermore, in actual operation, filter networks might be necessary at the receiver end to reduce the background noise from biological sources, man-made sources, or the specific "sea state". Appropriate reduction in the transmission bandwidth necessary to reduce these interferences to an acceptable level may further reduce intelligibility.

PURPOSE

The present investigation was conducted as part of a multi-phased study undertaken to evaluate several different aspects of "time-smearing" upon the overall intelligibility of speech. A detailed analysis of the intelligibility of spoken consonants in "time-smeared" speech would provide information about the relative effect of the smearing process on various consonant sounds and permit the selection of the more easily understood sound, thereby improving overall intelligibility. The information would be valuable in constructing special vocabularies for use in communicating by voice from one submarine to another when the speech signal is degraded by "time-smearing" distortion. Therefore, the purpose of the present investigation was to evaluate phoneme intelligibility of speech subjected to "time-smearing".

PROCEDURE

Recordings of a speech intelligibility test distorted by a computer to simulate the effects of "time-smearing" were presented to panels of listeners whose responses were analyzed for intelligibility of phonemes.

Recording of Stimulus Material. Recordings were made of the 50-item wordlists A through F of a Modified Rhyme Test (MRT) (House, et al, 1963)[1] using a high quality microphone and tape recording system. An adult male talker who was experienced in preparing intelligibility test stimuli read the wordlists in a general American dialect while the input to the recording system was monitored with a VU-meter to insure a constant recording level.

Computer Treatment. The acoustic transmission path through the ocean was simulated by computer to be a linear time-invariant channel. The channel was completely characterized by its impulse response. For the transmitted signal, s(t), the received signal, y(t), was the convolution of the transmitted signal, s(t), with the impulse response, h(t). A digital computer created the received signal, y(t), by utilizing a Fast Fourier Transform (FFT) technique to convolve digitized speech, s(t), with the digitized channel impulse response, h(t). The impulse response, h(t), was a sample function from Gaussian noise with a flat spectrum limited to the range 0-5 Kilohertz (kHz). The degree of "time-smearing" was determined by the duration of this impulse response in msec.

Organization of Speech Materials. The recorded wordlists A through F of the MRT were computer processed, organized into a complex set of stimulus tapes, and presented to groups of Ss at different combinations of low-pass filter and time-smear treatments. The overall organization of the experimental design was determined by a larger, more extensive study.

The four smear durations were 0, 50, 100 and 300 msec, and the eight filter settings were 400, 800, 1200, 1600, 2000, 2800 and 4000 Hertz (Hz) low-pass. Six MRT Wordlists were arranged in semi-random fashion to yield one recorded list for each of 32 filter-smear combinations. These 32 recordings were the basis of 8 "stimulus tapes", two of each of the four smear durations. In addition, two more recordings of MRT lists were produced: one to examine effects of digitizing and reassembling speech material with the computer, and the other to assess differences in listener-groups. These latter recordings were sectioned into five-word pieces which were alternately interchanged to produce two equated control lists.

For each of the four smear durations, one group of listeners was presented with a stimulus tape that contained half of the eight filter conditions and one of the two control conditions, while a second group of listeners was presented with a stimulus tape that contained the other half of the filter conditions and the other control conditions. In order to reduce the effects of order of presentation, the lists comprising the stimulus tapes were segmented into

ten-item sets and rearranged randomly. It was concluded that from the elevation of responses to the two control lists by the eight groups of listeners (two for each smear duration) that the groups performed very similarly in responding to the experimental material used in this study. Therefore, the data from the 320 subjects which made up the groups of listeners were combined to create a set of responses by one hypothetical group of 40 random listeners who heard one MRT list of 50 words for each of the 32 possible filter/smear conditions.

For the purpose of the present report, only that data pertaining to low-pass filter conditions of 1200 and 400 Hz for smear durations of 0, 50, 100 and 300 msec was studied.

Presentation of Speech Material to Listeners. The experiment was conducted in a group audio-testing room that was equipped with 50 matched TDH-39 earphones embedded in supra-aural cushions. The stimulus tapes were presented to 320 normal-hearing Navy enlisted men divided into eight groups of 40 each (one group for each of eight stimulus tapes). The words were presented monaurally at a level of 70 db SPL measured at the ear, using a Flat-Plate Acoustic Calibration Coupler. Listeners responded on specially constructed answer sheets (See Russotti and Duffy, 1969[2]) containing 6-word multiple choice blocks for the MRT word lists. Intelligibility was defined as the percent of correct responses.

RESULTS

Summarized results for the two low-pass filter conditions at the four time-smear durations are presented in Table 1. In the upper half of the table each entry is the average of the individual phoneme intelligibility scores for a filter/smear condition. The phoneme scores have been translated to the percent of correct responses out of the total number of presentations for the particular phoneme. Because of the use of percentages, the average scores in the upper half of Table 1 represent an intelligibility index that assumes each sound was presented an equal number of times in any complete 50-word MRT list. For comparison, the entries in the bottom half of Table 1 are the more conventional type of mean intelligibility scores. They were obtained by averaging scores across the 40 hypothetical listeners (described above in Organization of Speech Materials). In this latter situation, the actual number of presentations for each sound varied, and the score for each hypothetical listener was computed as the percentage of correct items out of the 50 for any filter/smear condition.

These two methods for assessing intelligibility produced very similar results. Except for the 4000 Hz/50 msec condition of filter/smear, mean differences in Table 1 were less than 2.8 percentage points for the two kinds of means. It may be noted that, with the same exception, intelligibility consistently drops as smear duration increases

TABLE 1

Mean percent of correct responses for different filter and smear-time conditions. Entries in the top half of the table are the averages of individual phoneme intelligibility. The bottom half of the table contains the mean percent of correct responses for 40 hypothetical subjects for a particular filter/smear condition according to the more conventional method of obtaining a single score per subject for a complete 50-word list of the MRT.

Low-Pass Condition	Smear-time in msec 0	50	100	300	
1200 Hz	59.5	45.3	41.0	40.2	Average of Individual Phoneme Intelligibility
4000 Hz	87.6	91.2	79.6	75.8	
1200 Hz	58.8	45.4	40.6	37.0	Conventional Mean Intelligibility
4000 Hz	85.6	84.0	77.4	76.2	

for both conditions of low-pass filtering. However, the reduction of intelligibility is much less for increased smear duration than for increased filtering. Incidentally, the results shown here for the 4000 Hz condition are nearly identical to the results obtained when these same recordings were time-smear-processed through the computer without filter conditions (Sachs et al., 1969)[3].

Table 2 presents the individual phoneme intelligibility scores in percent of correct responses as a function of smear duration for the 4000 Hz low-pass filter condition, and Table 3 presents similar information for the 1200 Hz low-pass filter condition. When smear duration is increased for the 4000 Hz condition, intelligibility decreases most for the voiceless stop-plosives [p, t] and [k]. Only two other sounds, the voiceless fricative [f] and the voiced stop-plosive [d] show a severe drop in intelligibility. Therefore, if a vocabulary is to be constructed for use where speech is affected by time-smearing, the phonemes [p, t, k, d] and [f] should be avoided if possible. Note that time-smear does not adversely affect the intelligibility of the semi-vowels; scores consistently remain in the 90's. These sounds appear to be highly resistant to distortion associated with time-smear, as do the voiced fricatives.

Results for the 1200 Hz filter condition presented in Table 3 show much greater variability than the results in

TABLE 2

Phoneme intelligibility scores in percent of correct responses as a function of smear time, for the 4000 Hertz low-pass filter condition.

Feature of Sound		SOUND	Smear Time in MSEC 0	50	100	300
FRICATIVES	VOICED	z	97.5	100.0	100.0	90.0
		v	87.5	87.5	92.5	87.5
		ð	85.0	---	52.5	85.0
	VOICELESS	f	89.2	82.5	70.8	61.6
		s	95.0	96.3	93.8	90.6
		θ	60.0	82.5	55.0	71.3
		ʃ	97.5	100.0	92.5	--
STOP-PLOSIVES	VOICED	b	81.7	83.3	70.0	76.3
		d	90.0	90.0	87.5	65.8
		g	91.7	76.3	95.8	90.8
	VOICELESS	p	81.7	90.0	55.8	57.5
		t	91.5	70.0	74.5	38.8
		k	80.5	100.0	68.0	55.0
NASALS		m	91.3	95.8	72.5	80.8
		n	90.0	93.8	85.6	96.9
		ŋ	85.0	97.5	80.0	--
SEMIVOWELS		r	97.5	97.5	97.5	95.0
		l	82.5	97.5	92.5	98.8
		w	98.8	97.5	96.3	95.0

TABLE 3

Phoneme intelligibility scores in percent of correct responses as a function of smear time for the 1200 Hertz low-pass filter condition.

Feature of Sound		SOUND	Smear Time in MSEC 0	50	100	300
FRICATIVES	VOICED	z	23.1	80.8	26.9	--
		v	11.5	42.3	3.8	11.5
		ð	--	7.7	--	--
	VOICELESS	f	61.5	30.8	46.2	34.6
		s	34.6	44.2	16.7	20.0
		θ	--	--	--	53.8
		ʃ	--	11.5	--	7.7
STOP-PLOSIVES	VOICED	b	82.7	19.2	43.3	79.5
		d	58.7	47.4	36.5	13.5
		g	78.8	11.5	55.8	75.0
	VOICELESS	p	54.8	50.0	31.7	30.8
		t	34.6	20.8	26.9	24.0
		k	74.0	32.3	52.9	25.0
NASALS		m	69.2	51.9	56.4	29.5
		n	66.3	47.4	26.9	35.6
		ŋ	19.2	84.6	11.5	46.2
SEMIVOWELS		r	100.0	80.8	50.0	88.5
		l	96.2	73.1	88.5	73.1
		w	100.0	75.0	92.3	65.4

Table 2. Consequently, it is more difficult to see the effects of increased smear duration. This is especially true for the fricatives and voiced plosives. However, as with the 4000 Hz filter condition, a consistent decrease in intelligibility occurs for voiceless plosives as smear duration is increased. Note that the semi-vowels no longer are immune to variations in smear duration when the signal is passed through the 1200 Hz low-pass filter. A comparison of Tables 2 and 3 reveals an apparent interaction between the semi-vowels and the 1200 Hz low-pass filter condition which is deleterious to the intelligibility of speech.

From analyses of Tables 2 and 3 we may conclude that voiceless plosives should be avoided whenever possible if time-smearing conditions similar to those studied here are imposed upon the speech signal. If there are no severe additional filtering restrictions, the semi-vowels [r, l] and [w] are the preferred phonemes for maximizing intelligibility. In fact, although affected by increased smear duration, the semi-vowels still produce greater intelligibility than do other sounds, when only frequencies below 1200 Hz are passed.

Phoneme intelligibility scores were used to compute Pearson Product Moment Correlation Coefficients (Table 4) among the different durations of time-smearing. Entries in the upper-right half of the table were calculated for the 1200 Hz low-pass filter condition from scores in Table 3, and those in the lower-left were calculated for the 4000 Hz low-pass filter condition from scores in Table 2. The lowest correlation for either filter treatment was between scores of the 0 and 50 msec smear durations. Under the 1200 Hz low-pass filter condition, the phoneme intelligibility for the 50 msec smear duration correlated poorly with that of the other three smear times (r = .01, .08, .12). Likewise, under the 4000 Hz filter condition, the r's of .34, .41, and .56 between phoneme intelligibility at 50 msec smear duration and phoneme intelligibility at the other smear durations are lower than most of the remaining correlation coefficients.* Only the r of .56 reached the 95 percent level of statistical confidence. Apparently, something happened to the pattern of phoneme intelligibility scores at the 50 msec smear duration so that they are not predictable from the scores obtained for the other smear durations tested. We may conclude from these data that phoneme intelligibility for the 50-msec duration cannot be predicted from phoneme intelligibility obtained under other time-smearing conditions. Conversely, those phoneme intelligibility scores obtained under 50 msec smear conditions are not good predictors of the scores obtained under other time-smearing conditions. In contrast, the statistically significant positive correlations among the 0, 100 and 300 msec smear conditions support the notion that the most intelligible phonemes under one smear condition generally will be the most intelligible phonemes under other conditions. The exception to this relationship occurs for the 0 and 300 msec smears when combined with 4000 Hz low-pass filtering. The low correlation between the data of the 50

The 0/300 msec time–smear correlation coefficient of r = .40 is an exception.

TABLE 4

Correlations between different smears for specific phoneme intelligibilities.

SMEAR CONDITION IN MSEC	SMEAR CONDITION IN MSEC: 0	50	100	300
0		.01	.78	.73
50	.34		.12	.08
100	.73	.41		.51
300	.40	.56	.63	

4000 Hertz Low-Pass Filtering

1200 Hertz Low-Pass Filtering

msec condition and other smear conditions, especially when subjected to 1200 Hz low-pass filtering, indicates the need for a separate, special vocabulary should time-smearing conditions similar to 50-msec duration be present.

SUMMARY AND CONCLUSIONS

One distortion of an acoustic signal which is caused by its transmission through the ocean is called "time-smearing". This study investigated the effect of time-smearing on the intelligibility of specific sounds of speech. Recorded tests of speech materials were treated by computer to simulate varied degrees of time-smearing. In addition, the speech was further degraded by either 4000 Hz or 1200 Hz low-pass filtering. Responses by 320 sailors who heard this speech material were analyzed for phoneme intelligibility. The reduced intelligibility produced by changing from 4000 Hz to 1200 Hz low-pass filters was much greater than the reductions attributable to variations in smear duration. As the duration of smearing increased, intelligibility dropped significantly. The voiceless stop-plosives [p, t] and [k] were the phonemes most severely affected by the time-smear distortion and consequently, should be avoided. On the other hand, the semi-vowels [r, l] and [w] were least affected and should be incorporated into any vocabulary designed to alleviate time-smearing distortion. Results in this report of the

intelligibility of specific speech sounds should be considered when constructing special vocabularies for use in transmitting signals through oceanic conditions where time-smearing effects are present.

REFERENCES

1. House, A. A., Williams, C., Hecker, M. H. L., and Kryter, K. D., Psychoacoustic Speech Tests: A Modified Rhyme Test. Electron Systems Division., USAF Sys. Com., Hanscom Field, Bedford, Mass., Technical Documentary Report No. ESD-TDR-63-403, 1963.

2. Russotti, J. S. and Duffy, J. F., An Evaluation of Three Methods for Unscrambling Helium Speech Produced at Depths of 800 and 1000 Feet. Naval Submarine Medical Research Laboratory, Report No. 602, October 1969.

3. Sachs, M. B., O'Brien, G. J., Sergeant, R. L. and Russotti, J. S., Speech Intelligibility in a Stationary Multipath Channel. Naval Submarine Medical Research Laboratory Joint Report with Naval Underwater Sound Laboratory, No. 574, March 1969.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Submarine Medical Center Submarine Medical Research Laboratory	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE
Specific Phoneme Intelligibility of Words Subjected to Time-Smearing Distortion

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Interim Report

5. AUTHOR(S) *(First name, middle initial, last name)*
Russell L. Sergeant, Joseph S. Russotti and Harry S. Cooker

6. REPORT DATE 25 January 1971	7a. TOTAL NO. OF PAGES 9	7b. NO. OF REFS 3

8a. CONTRACT OR GRANT NO. b. PROJECT NO. BuMed M4306.03-2020D NavShipsSystCom Task 5-23-18-08622 (USL 9-A-611-03-00) ONR/UCONN N00014-67-A-0197-0001	9a. ORIGINATOR'S REPORT NUMBER(S) SMRL Report No. 649 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Submarine Medical Center Box 600, Naval Submarine Base NL Groton, Connecticut 06340

13. ABSTRACT
One distortion to speech transmitted through the ocean is known as "time-smearing". In this study a detailed analysis was made of the intelligibility of spoken consonants in time-smeared speech. Recorded tests of speech materials were treated by computer to simulate varied degrees of time-smearing. In addition, the speech was further degraded by either 4000 or 1200 Hertz low-pass filtering. The responses of 320 sailors who heard this material were analyzed for phoneme intelligibility. Results show the reduction in intelligibility produced by changing the filter condition was much greater than reductions caused by altering smear duration. As the duration of the smearing increased, intelligibility dropped significantly. The voiceless stop-plosives [p,t] and [k] were the phonemes most severely affected by the time-smearing distortion while the semi-vowels [r,l] and [w] were least disturbed. Conclusion: the intelligibility of specific speech sounds should be considered when constructing special vocabularies for transmitting signals through the ocean when time-smearing effects are present.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0102-014-6600

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Speech Intelligibility						
Underwater Communication						
Special Vocabularies						
Phoneme Intelligibility						
Submarine Communication Systems						

DD FORM 1 NOV 65 1473 (BACK)

(PAGE 2)

UNCLASSIFIED

Security Classification